A Tag-based Network Evolution Mechanism for Online Communities

Caihong Sun
Information school, Renmin University of China, Beijing, 100872, China
chsun@ruc.edu.cn

Yan Xu
Georgia Institute of Technology, TSRB, 85 5th street NW, Atlanta, GA 30308,USA
yxu7@mail.gatech.edu

Xiaoping Yang
Information school, Renmin University of China, Beijing, 100872, China
yang@ruc.edu.cn

Abstract

An online community involves many interactions among members. When the interactions are frequent and important enough, they are usually explicitly shown as "friend links" among community members. In this paper, a tag-based network evolution model is proposed to simulate the growth and evolution of social networks of online communities. Our model is an agent-based model in which community members are simulated as agents who can decide whether to link to other agents as friends or not, according to their tag similarity. A tag is a value that represents an agent's total trait. Agents' traits could be their interests, locations or ages. We demonstrate the emergent properties of online communities generated by our tag-based evolution model, and compare these properties with those of other three network evolution models: BA model, small world network and random network model. Furthermore, two real-world online communities are studied empirically and compared with the online communities evolved from our model. Experimental results show that our evolution mechanism can facilitate the formation of a higher-quality community in which distances between members are smaller and members are more closely clustered.

1. Introduction

Online communities (i.e. Wikis, Blogs, Forums, etc) have become greater knowledge sharing resources in recent years. Online community or virtual community is a group of people communicating or interacting with each other by means of information technologies, typically the Internet, rather than in person [1]. In brief, an online community is a kind of computer-supported social network (CSSN) [2], in which members and their "friend links" form social networks. In such complex social networks, network structure plays an important role understanding higher-level issues, such as teasing out the prominent patterns in networks, tracing the flow of information, and discovering what effects these relations and networks have on people and organizations [3].

The "relations" we researched for online community are the explicit links listed on the home pages of community members. Of all the properties researched in social network analysis, we choose the following three, which are used in many previous works [4, 5, 6, 7, 8]:
(1) *Degree distribution*: Degree of a node (k) is the number of links it has. Degree distribution describes the probability distribution of degrees in a network.
(2) *Average path length* is the average of the shortest path between each pair of nodes. Average path length can describe how fast information can travel in a network.
(3) *Clustering coefficient* of a node is the proportion of links between the nodes within its neighborhood divided by the number of links that could possibly exist between them [8]. Clustering coefficient of a network is the average of all the nodes' clustering coefficients. It is used to describe how closely friends are clustered in a network.

Social networks are important in many aspects, but it is difficult to define and measure them, let alone their dynamic evolution over time. Erdos and Renyi in 1960 [18] proposed a random evolution model of network, Watts and Strogatz in 1998 [14], Barabasi and Albert in 1999 [19] proposed small world and scale-free network evolution models respectively. They provided new emerging simulation techniques to investigate the dynamics of a social network on the evolution of the network itself overtime. In this paper, we propose a tag-based model to simulate the growth and evolution of the network structures of online communities.

Holland first brought forward the concept of "tag" in 1995. A tag works like a flag that identifies one group of users from another. Holland assumed that arbitrary, evolving tags could facilitate selective interactions and thereby be helpful for aggregation and boundary formation [9]. From then on, many researchers conducted "tag" based simulation in different fields. In Riolo's work published on Nature, they used computer simulation methods and demonstrated that tag-based mechanism could lead to the emergence of cooperation even when the agents do not receive reciprocity and are unable to

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

observe or remember others' actions [10]. Tag-based mechanism is also applied to other different fields, such as "social rationality" [11], "cooperative peer-peer network" [12], and "Prisoners' dilemma game in small-world networks" [13].

In our tag-based mechanism, a tag is treated as the total trait of an agent in online communities. It is a total measure of characteristics of members, such as interests, locations or ages which are important factors for the formation of a social network structure. Our main idea is that members of online community tend to add those who have a similar total trait as their friends. In our model, members are simulated as agents who decide whether to link another agent as its friend or not, according to their tag similarity. The members of online communities usually learn from their friends, accordingly, in our model, a tag of an agent evolves to converge upon its friend's tags. Computer simulations show that the tag-based mechanism we proposed could facilitate the formation of a higher-quality community, in which friend links are more evenly distributed, distances between members are smaller and members are more closely clustered.

The rest of this paper is organized as follows: In Section 2, a tag-based mechanism is proposed to simulate the evolution of the social network structures of online communities, and four models are given based on tag-based evolution mechanism and preferential attachment mechanism. In Section 3, both simulation results and empirical studies are used to demonstrate the effectiveness of our tag based mechanism. Conclusions are drawn based on empirical studies and our simulation analysis. Section 4 concludes this paper and discusses future work.

2. Our Tag-based Network Evolution Model

In online communities, how people join the community and how they link to their friends are critical to the formation of their social network structures. The basic idea of our model is that effective network structure rises when agents tend to connect with people who share "similar enough" characteristics with them, while the community network is a directed graph. To test this idea, we propose a tag-based network evolution mechanism, and two models based on this mechanism given. Furthermore, two preferential attachment network evolution models based on the idea of BA model [7, 9] are simulated. Experimental results of these models are demonstrated and analyzed.

2.1. Tag-based mechanism

Every agent is assigned with a randomly generated tag and a threshold at the beginning. "Tag" (denoted as τ) is the total trait of an agent, which could be calculated from an agent's interests, locations or other observable traits. A tag is a real number and an agent's tag initially is generated from the uniform distribution with range of [0, 1]. Threshold (denoted as T) is a real number, which is the degree of tolerance by which an agent chooses his friends. Agents have their own private degree. An agent's threshold is generated from uniform distribution with range of [0, 0.5]. Our tag-base mechanism is that an agent randomly selects some agents as its friend candidates, but only those whose tag's difference are smaller than its difference tolerance threshold will become its friends.

Model 1

In Model 1, the network evolution of an online community follows three rules:

Rule 1. Growth of agents: Starting with a small random network (with m_0 (=10) node, and p=0.1), and at each time step, we add one new agent. The tag and threshold of the new agent is randomly generated. All agents follow Rule 2 to add their links.

Rule 2. Growth of links: At each time step, the new agent has chance to select one node as its own "friend". The selection includes: first, the agent A randomly gets k_0 other agents in the network as friend candidates; second, for each candidate agent B, agent A compares its own tag with B, if the tag difference is no more than its own threshold ($|\tau_A - \tau_B| \le T_A$), then agent A adds agent B as its friend, and the selection procedure ends.

Rule 3. Learning from friend's tag: At every time step, each agent adapts its own tag according to the tag of its new friend, with a learning rate of c ($0 \le c \le 1$), as shown in Equation (1). Where t is the time step. If $\tau_A(t+1)$ is larger than 1 or smaller than 0, it is set to 1 or 0.

$$\tau_A(t+1) = \tau_A(t) + c \times (\tau_B(t) - \tau_A(t)) \qquad (1)$$

Rule 3 is based on the assumption that people tend to have similar interests or traits with his friends.

Model 2

Besides the three rules in model 1, a wiring rule is added as: in every q step, l random agents in the existing network get an opportunity to add a new friend, and their friend selection process follows Rule 2 and Rule 3 in Model 1. The Wiring rule simulates the fact that agents tend to make new friends after they join the community. In addition, the tag learning rule (Rule 3) in Model 1 is changed to Equation (2). Where $\Gamma(A)$ is a set of the friends of A.

$$\tau_A(t+1) = \tau_A(t) + c \times \left(\frac{\sum_{i \in \Gamma(A)} \tau_i(t)}{|\Gamma(A)|} - \tau_A(t)\right) \qquad (2)$$

0-7695-2875-9/07 $25.00 © 2007 IEEE

2.2. Preferential attachment mechanism

Model 3

Model 3 follows a network evolution mechanism proposed by Barabasi and Albert [19]. The online community network growth follows two rules:

Rule 1. Growth of agents: Starting with a small random network (with m_0 (=10) node, and p=0.1), and at each time step, we add a new agent. All agents follow Rule 2 to add their links.

Rule 2. Growth of links (preference attachment): At every time step, the new agent selects one of the existing agents as its friend with a certain probability. The probability is proportional to the number of links that the existing agent already has.

The number of links could be seen as an agent's reputation in a community. In Model 3, a new agent tends to choose agents who have higher reputation as its friends.

Model 4

The online community network growth rules are same with model 3 except that in Rule 2, the probability is proportional to tag similarity between the existing agent and the new agent.

3. Experimental Analysis &Empirical Studies

3.1 Experimental analysis

Experiment 1: *What is the degree distribution of the network generated by each model?*

In this experiment, the simulations of the network growth of each model are executed. The number of agents in each network is 5000 (i.e. n=5000). The experimental results are shown in Figure 1.

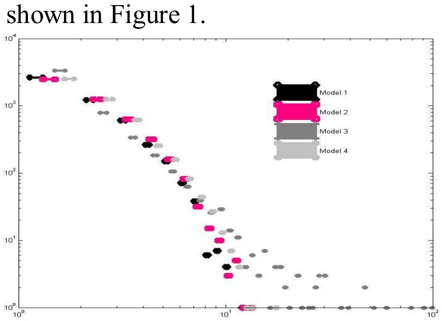

Figure 1. In-degree distribution (Log-Log plot)

Note: In experiment 1, $q = 10$, and $l = 1$ in Model 2. It means the wiring rate is very low.

Model 3 demonstrates power law, which is consistent with BA model [7, 19]. The degree distribution of the tag-related models do not follow the feature of power law, but looks more like a parabola as shown in Figure 1. In a scale-free network, there exists hub agents, but in our tag-based model, there are no big "hub" agents, and no agents have been linked by more than 15 other agents.

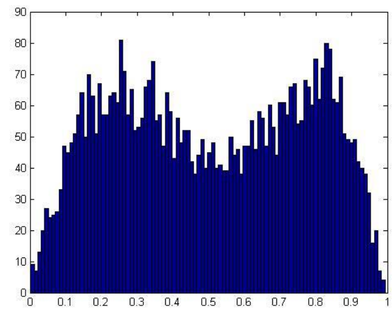

Figure 2. Histogram of tags (n=5000, Model 2)

In contrast to uniform distribution, tags are clustered into some peaks, shown in Figure 2.

Experiment 2: *When number of agents is fixed, how clustering coefficient (CC) changes with average degree ($\bar{k}$)?*

In this experiment, we set the number of agents to a fixed number (n=10,000), and use Model 2 as the evolving model. By controlling the parameter of wiring rate, we obtain networks with different $\bar{k}$ (average degree).

Usually clustering coefficient is positively related to average of a network's degree ($\bar{k}$). However, in our tag-based Model 2, clustering coefficient is a stable value, irrelevant with the increase of $\bar{k}$.

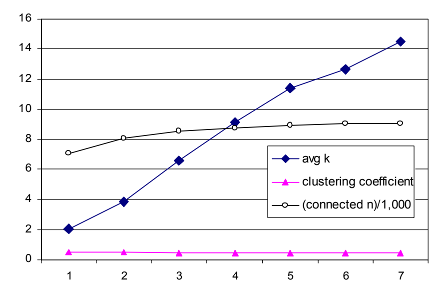

Figure 3. Clustering coefficient and Average Degree

In Figure 3 (n=10,000), the number of connected agents grows slowly with the increase of average degree. However, clustering coefficient stays around 0.45-0.47 while $\bar{k}$ changes from 2 to 14.5.

This result shows that when agents follow tag-based mechanism, no matter whether the users are lazy to make friends or they prefer to make more friends, they all share the high clustering of their friends. Agents are tightly connected. This is consistent with Kumar's assumption [4]: similar interests/age/location can cause the high clustering of "Friendship". In addition, threshold is the key parameter, it influences $\bar{k}$ when other variants are fixed: the larger the threshold is, the larger $\bar{k}$ is.

3.2 Empirical Studies

Two online communities are studied as social networks in this section. One is Douban.com, which is a website where people communicate one another and recommend books, music and movies. One is a web log website, on which users can read and write their digital "diaries". We use SiteX to represent the web log website because of the agreement with the company of sharing its data. Both sites are popular in web according to the traffic data flows reported by Alexa.com. On both sites, members own a personal home page, where they explicitly add the links to their friends' home page. We study the two online communities as friendship network.

The data from Douban.com are crawled by a breadth first search with the depth of 15. The data for SiteX are obtained from the service provider, which consist of the users created in a period of time (in one year). Some of the properties of the two samples are described in Table 1.

Table 1. Descriptions of the two samples

Community Name	Number of Nodes	Number of Links	Average node degree	Percentage of mutual links
Douban.com	13,949	64,887	4.65	49.2%
SiteX	34,527	23,098	0.67	24.7%

Experimental results show that the degree distribution of two online communities follows power-law distribution: $p(k) \propto k^{\gamma}$, where k represents node's degree. The Coefficient γ and R^2 are shown in Table 2.

Table 2. γ and R^2 of two online communities

	Douban.com		SiteX	
	γ	R^2	γ	R^2
In-degree	1.71	0.90	4.05	0.94
Out-degree	1.85	0.93	2.59	0.96

Experiment 3: *How are the clustering coefficient and average path length different from empirical data and those of other models?*

Clustering coefficient is the measurement for whether the friends are closely or loosely clustered. As we can see in Table 3, the clustering coefficient of Physics coauthor network is higher than computer-supported social network (such as email, LiveJournal, douban.com etc.), the former is 0.45 and the latter is among 0.1-0.2. Although our model depicts online community, the clustering coefficient is as high as real-world social network. It is 100 times of that of BA model, 400 times of Random network model. This feature shows the emergence of closer clusters by applying tag-based mechanism.

Average path length is a measurement for the time speed of information or innovation diffusion. Tag-based model also works well on this property. The average path length is only 1/3 of random network. The comparison results are shown in Table 3.

Table 3. Comparisons of tag-based model, BA model, random network and real-world networks

Name	Average degree	Total number of the network	Clustering Coefficient	Average path length
Tag-based Model 2	3.86	8040	0.476	1.89
BA model	5	8000	0.004	4.86
Random network	4	8040	0.001	6.61
E-mail message*	1.44	59,912	0.16	4.95
Physics coauthorship*	9.27	52,909	0.45	6.19
LiveJournal*	14	1.3 million (sampling 25,000)	0.2	-
Blog.onet.pl*	0.81	150,000	-	7.6
Douban.com	4.65	70,000 (Sampling 13,949)	0.145	6.78
SiteX	0.67	2 million (sampling 34,527)	-	9.55

*THE DATA OF REAL-WORLD NETWORKS ARE EXTRACTED FROM NEWMAN'S SUMMARY [17].

3.3 Summary

As demonstrated in our experiments above, our tag-based mechanism (Model 2) can generate a higher-quality community in which friend links are more evenly distributed, distances between members are smaller and members are more closely clustered.

There are no top-down controls or policies included in our tag-based mechanism since the sociability of an online community can not be controlled. Compared to the BA model [7], in which newly added agents tend to connect with those existing agents having many friends, the tag-based mechanism depends on an agent's own tag. Agents are paired at random, and one agent can decide whether to form a link to the other agent or not, based on the similarity of their tags.

But what will happen if Rule 3 (learning from friends' tags) is removed from our model? Our simulation results show that the shape of degree distribution does not change, but the clustering coefficient decreases to 0.35 (N=10000, $\bar{k}$=14), the average path length sharply grows up to 89. This means that the members are fragmented into many small groups. A tag's evolvement (Rule 3) is quite important in our model, and this also fits well with the fact that people tend to have similar interests with their friends.

From our empirical studies, we know that the networks of the two online communities are sparse, indicating that most people do not maintain many friends in online communities. Meanwhile, the degree distribution is found to be scale-free, showing that a very small fraction of the community members are "hubs" of the network. These "hub" members link to or are linked by a lot of other members. But "hubs" may cause some problems: if these hubs become inactive or be attacked, the connectivity of the network will be greatly harmed [15]. If the links in the online community were more evenly distributed, it would be more robust to resist the intended attacks on hub members, and more interactivity between members would be provided. In addition, compared to other social networks, such as the film actor network [8], the clustering coefficient of the two online communities is relatively small. A higher clustering coefficient indicates a closer friend circle, which is the basis for providing more social support. And our simulation results show that by applying the tag-based evolving mechanism, a higher-quality online community could be formed.

4. Conclusions and Future work

In this paper, a tag-based network evolution mechanism is presented. As demonstrated in our experiments in section 3, our tag-based models can generate a higher-quality community in which friend links are more evenly distributed, distances between members are smaller and members are more closely clustered.

Computer simulation results of our proposed mechanism show that our models demonstrate interesting properties. (1) a larger and more stable clustering coefficient indicates that the tag-based mechanism is helpful to form closely connected clusters; (2) more evenly degree distribution shows the potential of introducing more agents with a medium number of friends, rather than a very small number of densely connected hubs, which would be the basis for more mutual interactivity; (3) the smaller average characteristic path length shows that tag-based model can diffuse information and innovation more quickly.

In addition, the tag-based mechanism we proposed is bottom-up; no central control is required to form all these good social structures. Each agent only needs to compare another agent's tag with its own, degree or tag information about every agent in the network are not required.

Our future work includes: (1) introducing tag with multiple dimensions; (2) studying the emergent properties of the mechanism which combines tag-based mechanism with preferential attachment mechanism; and (3) considering the life cycle of the agents. In this paper, we only use a value (tag) to indicate the total trait; in the future, we will study the effects of multi-dimension tag which presents multiple traits of agents. We will examine the effects on friend networks by breaking some friend links and introducing agents' life span.

References

[1] http://www.answers.com/topic/virtual-community
[2] B.Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia and C. Haythornthwaite, Computer Network as Social Networks: Collaborative Work, Telework, and Virtual Community, Annual Review of Sociology, Vol. 22 (1996), 213-238.
[3] L. Garton,, C. Haythornthwaite and B. Wellman, Studying Online Social Networks. Journal of Computer Mediated Communication 3 (1) (1997).
[4] R. Kumar, J. Novak, P. Raghavan and A. Tomkins, Structure and Evolution of Blogspace, Communications of the ACM, December 2004/Vol. 47, No. 12, p35-p39.
[5] W. Bachnik, S. Szymczyk, P. Leszczynski, R. Podsiadlo, E. Rymszewicz, L. Kurylo, D. Makowiec and B. Bykowska, Quantitive and sociological analysis of blog networks, Acta Physica Polonica, Vol. 36 (2005), No.10, 3179-3190.
[6] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. ISBN: 0521387078, Cambridge University Press (1994) 4-9.
[7] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74, (2002), 47-97.
[8] D.J. Watts, Small World, the Dynamics of Networks between Order and Randomness. Princeton University Press (1999), chapter 2,3.
[9] J. Holland, Hidden Order: How Adaptation Builds Complexity (Addison Wesley, Reading, MA, 1995).
[10] R. Riolo, R. Axelrod, and M.D. Cohen, Evolution of cooperation without reciprocity. Nature 414 (22 Nov 2001), 441-3.
[11] D. Hales, and B. Edmonds, Evolving social rationality for MAS using "tags", International Conference on Autonomous Agents archive, Proceedings of the second international joint conference on Autonomous agents and multi-agent systems, 2003, Pages: 497 – 503.
[12] D. Hales, From selfish nodes to cooperative networks-emergent link-based incentives in peer-to-peer networks, Peer-to-Peer Computing, 2004. Proceedings. Proceedings. Fourth International Conference on, p151- 158.
[13] B.J. Kim, A. Trusina, P. Holme, P. Minnhagen, J.S. Chung and M.Y. Choi, Dynamic instabilities induced by asymmetric influence: Prisoners' dilemma game in small-world networks, Phys. Rev. E 66, 021907 (2002).
[14] D.J. Watts and S.H. Strogatz, Collective dynamics of 'small-world' networks. Nature 393, 440--442 (4 June 1998).
[15] R. Albert, H. Jeong and A.L. Barabási, Error and attack tolerance in complex networks, Nature 406, (2000), 378.
[16] R.I.M. Dunbar, Neocortex size as a constraint on group size in primates, Journal of Human Evolution 22, (1992) 469-493.
[17] M. E. J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): (2003),167-256.
[18] P. Erdos and A. Renyi, On the evolution of random graphs[J]. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, 5: 17-61.
[19] A. L. Barabaasi and R. Albert, Emergence of scaling in random networks. Science, 1999, 286: 509-512.